DOCUMENT RESUME

ED 217 406 CS 006 729

AUTHOR Anderson, Richard C.; Freebody, Peter
TITLE Reading Comprehension and the Assessment and Acquisition of Word Knowledge. Technical Report No. 249.
INSTITUTION Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Jun 82
CONTRACT 400-76-0116
NOTE 51p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Aptitude; *Cognitive Processes; *Evaluation Methods; Grade 5; Intermediate Grades; Language Research; Language Skills; *Reading Comprehension; *Reading Research; Semantics; *Vocabulary Development; Vocabulary Skills; *Word Recognition

ABSTRACT

The "yes/no" method of vocabulary assessment requires students to indicate words they know from among a list of words and nonwords. Preliminary evidence gained from a study involving fifth grade students indicates that the method is superior in many ways to the multiple choice method of assessment. Analysis of "false alarms," cases in which children say they know the meanings of nonwords, reveals that good readers aggressively apply morphological rules to hypothecate meanings for unfamiliar terms, whereas poor readers engage in phonemic experimentation with unfamiliar items to transform them into common words. A review of the literature shows that vocabulary difficulty is a factor in text comprehension, but that it is not as important as studies of readability have suggested. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED217406

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

CENTER FOR THE STUDY OF READING

Technical Report No. 249

READING COMPREHENSION AND THE ASSESSMENT AND ACQUISITION OF WORD KNOWLEDGE

Richard C. Anderson
University of Illinois at Urbana-Champaign

Peter Freebody
University of New England

June 1982

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02238

The research reported herein was supported in part by the National Institute of Education under Contract No. US-NIE-C-400-76-0116. The authors gratefully acknowledge the contributions that have been made to this work by Michael Levine and William Nagy. To appear in B. Hutson (Ed.), Advances in Reading/Language Research, A Research Annual. JAI Press Inc., Greenwich, Connecticut.

CS006729

EDITORIAL BOARD

William Nagy and Stephen Wilhite
Co-Editors

Harry Blanchard
Charlotte Blomeyer
Nancy Bryant
Larry Colker
Avon Crismore
Meg Gallagher
Anne Hay
Asghar Iran-Nejad
Margi Laff
Jill LaZansky
Cindy Steinberg
Terry Turner
Janet Williams
Paul Wilson

Abstract

An evaluation is presented of a method of vocabulary assessment, called the yes/no method, in which students indicate the words they know from among a list of words and nonwords. Preliminary evidence indicates that the yes/no method is much better in several respects than the multiple choice method. Analysis of "false alarms," cases in which children say they know the meanings of nonwords, reveals that good readers aggressively apply morphological rules to hypothecate meanings for unfamiliar items whereas poor readers engage in phonemic experimentation with unfamiliar items to transform them into common words. Studies are summarized that show that vocabulary difficulty is a factor in text comprehension, but not as important a one as studies of readability suggest.

Reading Comprehension and the Assessment and Acquisition of Word Knowledge

Our intuitions about how our native language works are strong: Even subtle violations of grammar or conventions of usage ring loudly in our ears; we make rapid and usually accurate predictions about the content and interest value of a speech on the basis of the speaker's first few sentences; we also make generally appropriate estimates of the intellectual abilities of a speaker or writer based on a similarly small sample of language. One of our stronger intuitions is that familiarity with the words used in an utterance is a reliable touchstone from which we can infer how manageable we will find the meaning of an utterance. That is, from knowledge of the vocabulary, we infer the accessibility of the message.

It is true, of course, that exotic words can be used to dress up a banal message. Deliberate pomposity in language is not uncommon. The intuition that there is a close relation between familiarity of the vocabulary and the difficulty of the conceptual content in a message is rattled frequently by social scientists. Consider this piece from one of the major social theorists of this century:

> *The problem of order, and thus of the nature of the integration of stable systems of social interaction, that is, of social structure, focuses on the integration of the motivation of actors with the normative cultural standards which integrate the action system, in our context inter-personally. These standards are patterns of value-orientation, and as such are a particularly crucial part of the cultural tradition of the social system.* (Parsons, 1951, p. 37)

This insight was translated by another important sociologist as:

> *People often share standards and expect one another to stick to them. In so far as they do, their society may be orderly.* (Mills, 1970, p. 36)

Quite a straightforward suggestion, but daunting in its original form.

More interesting and less common perhaps is the opposite case. It sometimes happens that our familiarity with the words bears little relation to the ease with which we can construct a meaning. There is something eerie about encountering a string of highly frequent, "easy" words in equally simple grammatical forms, and finding yourself unable to construct a meaning for the discourse.

Examine the following extract. It is the opening statement from a famous work on logic and philosophy.

> *The world is all that is the case.*
> *The world is the totality of facts, not of things.*
> *The world is determined by the facts, and by their being all the facts.*
> *For the totality of facts determines what is the case, and also whatever is not the case.*
> *The facts in logical space are the world.*
> *The world divides into facts.*
> *Each item can be the case or not the case while everything else remains the same.*
> (Wittgenstein, 1961, p. 7)

Words such as *world*, *case*, *facts*, *things*, and *items* are all quite familiar to us, but we nonetheless are left feeling that we have not quite penetrated the nebula of the author's communicative intentions.

The intuition that familiarity with individual words is a useful predictor of the effort needed to understand a piece of discourse is a

sound one, despite occasional slips of the kind illustrated above. This is reflected by the fact that every readability formula gives a heavy weight to vocabulary familiarity. Moreover, the breadth of a person's vocabulary has been recognized for some time as a very good predictor of that person's general intelligence (Terman, 1918) and reading comprehension ability (Thorndike, 1973), though, it should be added, it is far from clear why this is true.

Estimating vocabulary size has been a perennial concern of educational researchers. However, as we have shown elsewhere (Anderson & Freebody, 1981), estimates of the total word knowledge of individuals at various ages have fluctuated wildly. Comparison of estimates of vocabulary size indicates large discrepancies, by as much as a factor of 10. In the face of this uncertainty in the research literature, we find surprising the conviction voiced by language psychologists and reading experts that children acquire many word meanings with great ease and rapidity, at a rate which could not be accounted for by their exposure to formal instruction. The eminent psychologist, George Miller, for instance, recently claimed that the "best figures available" showed that children of average intelligence levels "learn new words at a rate of more than 20 per day" (Miller, 1978, p. 1003). Obviously schools do not directly teach 20 new words every day. Several reading educators, apparently under the influence of those same "best figures," have concluded that teachers ought not concentrate too heavily on instruction in word meanings, since, if the figures are accepted, apparently children learn most words on their own.

Important educational policy decisions hinge on having accurate information about how many words children of different ages know and how they came to know these words. If the year to year growth in vocabulary for the average child is as large as some figures suggest, then the best advice to teachers would be to help children become independent word learners, since direct vocabulary instruction could make only a pitifully small contribution. On the other hand, if typical year to year changes in vocabulary size are small, direct vocabulary instruction might be a viable practice.

The disturbing discrepancies in estimates of vocabulary size seem to have arisen for two reasons. First, there has been considerable variation in the operational definition of a "word" in English. Usually, definitions are dictionary based. The larger the dictionary the larger the estimate of vocabulary size. Also important are such questions as whether proper names, acronyms, technical terms, archaic words, slang, inflections, derivatives, and compounds will count as separate words. Researchers have adopted different approaches to these questions, with predictably different results.

Second, different methods of assessing word knowledge have led to different estimates of word knowledge. By far the most common format is the multiple choice procedure. We have argued (Anderson & Freebody, 1981) that there are two good reasons for questioning the validity of the multiple choice procedure as a measure of breadth of vocabulary knowledge. First, the distractors in a test item strongly influence performance. Second, test taking strategy is inevitably a factor in performance on multiple choice tests. This serves to disadvantage young

children and, perhaps, some older children, because they do not systematically consider all of the options. We will summarize later in this chapter data we have collected that calls into serious question the validity of the multiple choice method of measuring vocabulary knowledge.

By way of introduction, then, we hope that we have shown that vocabulary knowledge ought to be an important construct in models of cognitive functioning generally, and in models of reading comprehension in particular. We hope also that we have convinced you that there are problems in the area of vocabulary assessment. Gross discrepancies in estimates of word knowledge and fundamental uncertainties about modes of selecting word samples and procedures for testing knowledge point to the need for both conceptual and empirical clarification of several related questions: How can we assess word knowledge validly? How can we estimate the total number of words a person knows? How is new vocabulary acquired? What is the nature of individual differences in vocabulary knowledge? and What role does vocabulary knowledge play in reading comprehension? The answers we offer here to these questions will in part be based on data we have collected, in part on extrapolations from our data, in part on impressions gained while asking children the meanings of words, and occasionally on just plain speculation. Our overall goal in this paper is to stimulate thought and research and to offer to reading educators a procedure for assessing word knowledge and a way of thinking about the role of vocabulary which they might find, at least, interesting, and perhaps even useful.

The Assessment of Knowledge of Word Meanings

In this section we will present a nontechnical discussion of research we have undertaken to develop a better measure of vocabulary knowledge. The goals of our initial studies were to examine the efficacy of a method with minimal response and strategic demands and to compare the validity of such a measure with that of the most popular format, the multiple choice test. In order to allow the multiple choice test its strongest possible showing, we selected the vocabulary subscale of the Stanford Achievement Test (1973). Presumably these items and their distractors have been thoroughly analyzed. Presumably the items included in the test are neither too easy nor too difficult and have good discriminating power. The test as a whole is highly reliable and correlates highly with intelligence tests and other achievement tests.

We focused our study on fifth grade students. All the items at the fifth grade level of the Stanford vocabulary scale were used, and about one third of the items from the two levels above and the two levels below the level appropriate to fifth grade were randomly selected. This procedure yielded 195 multiple choice items, ranging in intended level from second to about ninth grade.

We are attracted to the simple yes/no method of vocabulary assessment, in which the student indicates by a check (or the press of a button or something equally simple) whether or not he or she knows the meaning of a word. The great a priori appeal of the method is that it strips away irrelevant task demands that may make it difficult for young readers and poor readers to show what they know. Performance on multiple choice items depends not only on whether the examinee knows the word being

tested, but also the nature of the distractors. Sometimes determining the right answer will require the examinee to know several other words as hard or harder than the tested word. Moreover, test taking strategies are a factor in performance on multiple choice tests. Young and underachieving examinees are less likely to possess these strategies--less likely, for instance, to consider all of the options rather than pick the first that strikes their fancy.

On the other hand, any approach to assessing vocabulary knowledge that requires freely composed answers will stress ability at exposition and, in the case of written answers, may depend in part on spelling or even penmanship. Evaluating freely composed answers is costly and involves difficult and somewhat arbitrary scoring decisions. The approach makes inefficient use of examiner and examinee time.

In contrast, the yes/no test would appear to minimize extraneous demands for strategic knowledge or ability in self-expression. The one great question about the yes/no method has been obvious since the early days of vocabulary testing (cf. Sims, 1929): What is to prevent people from overstating their vocabulary knowledge, checking "yes" for words they do not actually know?

To solve the problem of people using too lenient a standard in judging whether a word is known, we have devised a version of the yes/no task that includes like-English nonwords among the real words. It stands to reason that persons who indicate they "know" the meaning of very many nonwords are using too slack a standard.

Mixing words and nonwords in a vocabulary test is a variant of a laboratory procedure called the "lexical decision" task. We are not the

first to think of using the procedure to assess vocabulary knowledge (Zimmerman, Broder, Shaughnessy, & Underwood, 1977).

Determining precisely the right adjustment to make in order to correct for an individual's tendency to overestimate the number of words he or she knows is a knotty problem. In collaboration with Michael Levine, we are working on what we hope will turn out to be an elegant solution using latent trait theory. However, this work is not completed yet, so for the purposes of this paper we shall rely on a simple approach resembling the one educators have traditionally used to correct multiple and true/false tests for guessing. We have good reason to believe that this approach is satisfactory for most practical purposes.

Following conventional terminology, let us say that a student has scored a "hit" when he indicates that he knows the meaning of a real word but a "false alarm" when he says he knows the meaning of a nonword. The proportion of words truly known, $P(K)$, is estimated by the following simple formula:

$$P(K) = \frac{P(H) - P(FA)}{1 - P(FA)}$$

Consider two students who both say yes to 70% of the real words. One student has also said yes to 30% of the nonsense words, while the other has said yes to only 5% of the nonsense words. According to the formula above, the former student knows 57% of the words whereas the latter student knows 68%, which matches one's intuition that the former student was guessing more often. Technically speaking the formula provides a "high threshold" correction, since it is based on the assumption that when an examinee says yes he or she either knows the item perfectly or has made a blind guess.

The yes/no procedure was evaluated and compared to the multiple choice method in a study in which 120 fifth graders participated. The children completed a multiple choice test, consisting of the 195 English words as previously described, and a yes/no test involving the same 195 words. The yes/no test also included 131 nonwords. We made up the nonwords by changing one or two letters in real words (e.g. flirt became flort and perfume became porfame) and by forming unconventional base plus affix combinations (e.g. observement, adjustion) which we will henceforth call pseudo-derivatives.

One advantage for the yes/no format was immediately obvious. The children completed over three times as many yes/no items, covering over twice as many words, in a given period of time as they did multiple choice items. Machine scorable answer sheets were used for both tests. The relative time advantage of the yes/no probably would have been even greater if the children had been answering directly in the test booklet or taking the test at a computer terminal.

The correlation between multiple choice scores and corrected yes/no scores was .84. Whereas this is a strong relationship, it is not as strong as might be expected considering that the same 195 words are assessed. The two tests were administered one week apart. We suspect that the value of .84 is considerably below the one-week test-retest reliability of either measure. Since the two tests do not measure exactly the same thing, the question that naturally arises is which one gives the most valid assessment of vocabulary knowledge.

The sense of valid that will be used here is that a person's test score ought to indicate the proportion of words he or she actually knows

and that, alternatively, the average score of a group of people on a certain word ought to indicate the proportion in the group that actually know this word. To compare the validity of the yes/no and multiple choice tests, all of the fifth graders were interviewed about the meanings of a set of 40 words on which the two tests gave discrepant results. The set included 20 words that substantially more students claimed to know on the yes/no test than got correct on the multiple choice test and 20 words for which the reverse was true.

The children read each of the 40 words, had his decoding corrected if necessary, and then was asked what the word meant. The children were asked to define the word or, if they could not do that, to use the word in a sentence. If a child could neither define a word nor use it in a sentence, he or she was probed with questions such as "Can you tell me anything about it?" and, "What does it make you think of?" The experimenter played an active, Socratic role attempting to get the children to tell all they knew and asking questions to clarify ambiguous answers. The interview protocols were scored according to three different criteria: strict (the child could give an adult-like definition); moderate (the child could either define the word, or use it in a sentence that indicated knowledge of its meaning); lenient (the child met either of the first two criteria or produced an association that suggested knowledge of at least one distinction conveyed by the word).

For the 40 words, the correlations between the proportion of children who indicated on the yes/no test that they knew the meanings of the words and the proportion whose interview answers met the strict, moderate, and lenient criteria were .85, .89, and .92, respectively. The correlations

between the proportions of children whose interview answers met the three criteria and the proportion who got correct answers on the multiple test were .45, .43, and .45. This is a dramatic advantage for the yes/no test. Indeed, when the average proportion of hits for each word was corrected for average false alarm rate, the slope of the regression line predicting the proportion of children meeting the lenient interview criterion approached 1 and the intercept approached zero. For the multiple choice proportion, corrected for guessing, the slope of the regression line was much flatter, and there was a greater amount of fluctuation around that line; that is, the prediction was poor.

Some examples will illustrate the differences in performance. For the word manage, 72% of the students could give an adequate definition, 92% could define it or use it in a sentence satisfactorily, and 97% could define it, use it in a sentence, or give some semantically relevant information about it. On the yes/no test 96% said they knew manage, but only 28% got the multiple choice item correct. Here is that item:

If you manage on your allowance, you -

1. spend it
2. save it
3. get along
4. waste it

Many of the students selected the first choice. It is not only a plausible response, given the unimpressive amount of allowance most fifth graders receive, but it is in the first position. This gives it an advantage, since some children tend not to examine fully all the distractors, but will often choose the first or second one if it makes acceptable sense.

This tendency may have affected performance on the word apology, which only 56% of students claimed to know, but which 77% of them got right in its multiple choice format. The relevant multiple choice item is:

Words saying you are sorry are -

1. an apology 2. a defense 3. a pardon

The early appearance of the correct answer may have accounted for the enhanced performance.

Another case in which a word evidently was not well known but in which the distractors may have helped in the multiple choice test was judicious. About 19% of the students said yes to it on the yes/no test while 51% of the students got it correct in its multiple choice form. On the interview test, 2% could define it, 3% could use it in a sentence, and 24% could give some suitable association. The item is:

A judicious decision is made -

1. quickly
2. wisely
3. foolishly
4. cleverly

The association of the first three letters of the word with the word judge may have led students to the second option, or maybe students are sensitive to the fact that decisions are more often called wise than quick, foolish, or clever.

From examining our data, we have developed the generalization that when the word tested in a standardized multiple choice item is difficult something about the item will tend to give away the correct answer, whereas when an easy word is tested the item will tend to lead the student away from the correct answer. An objective measure of a word's difficulty is

its frequency of usage. The best measure of frequency is the frequency of the morphological "family" of which the word is a member. A morphological family consists of a basic word and all of its inflections and semantically transparent derivatives and compounds. Nagy and Anderson (Note 1) have presented a thorough discussion of the criteria for determining family membership and have provided estimates of the number of word families in printed school English. For the 195 words used in the present study, family frequency correlated .70 with yes/no proportion but only .51 with multiple choice proportion.

That performance on a standardized multiple choice test should bear only a modest relationship to a measure of intrinsic difficulty is not surprising. As one of us once put it, standard item analysis procedures "torture validity" (Anderson, 1972). When an item analysis shows that a question is "too easy" it will be thrown out. Thus, when the item is inherently easy, it will be kept only if it contains an irrelevant obstacle to comprehension. Conversely, a standard item analysis will cause an intrinsically difficult item to be rejected unless something about the item tends to give away the correct answer.

Our early indications are, then, that a person's score on a yes/no vocabulary test, suitably adjusted to discount any tendency to overestimate vocabulary knowledge, is an excellent indicator of the number of words this person truly knows. Several caveats are necessary, however. First, a yes/no test could not determine whether a person knew one of the particular meanings of a polysemous word, since presumably the person would say <u>yes</u> if he or she knew any of its meanings.

Second, a yes/no test is unsuitable for evaluating the effects of direct vocabulary instruction, since students will be able to recognize the words taught as familiar and say yes even though they don't know their meanings. Indeed, the possibility that people could answer yes/no items on the basis of familiarity, rather than knowledge of meanings, is a possible general problem with the yes/no test which we are currently evaluating.

Third, though the results summarized here indicate that a yes/no test provides a much better measure of whether the examinees know the meanings of the words tested than a standardized multiple choice test, the yes/no test may nonetheless have lower "reliability" and "predictive validity." The basis for this caveat is that successful performance on a multiple choice vocabulary test requires, in addition to knowledge of word meanings, reasoning, planful use of working memory to hold response options in mind, and sensitivity to the subtle nuances of language use in cultured, mainstream circles. This skill and knowledge is possessed in fuller measure by students of high ability or high socioeconomic status, and thus contributes to apparent reliability and predictive validity. The role of extraneous factors is exacerbated in performance on a standardized multiple choice because the test maker uses discriminating power as a criterion for including or excluding items.

Individual Patterns of Performance

An analysis of false alarms revealed a fascinating difference in the performance of high- and low-ability fifth graders. Table 1 shows the most frequent false alarms of the children who fell in the top and

bottom quartiles, based on total adjusted yes/no score, among the 120 fifth graders who participated in the study.

Insert Table 1 about here.

The first thing to note is that almost all of the false alarms of the high-ability children are pseudo-derivatives. The error rate is extraordinarily high on some of these items considering that the children are well above average in reading ability, that their average false alarm rate is only 6.4% and that on 65 of the 131 nonwords not one of these children false alarmed. On a few of the pseudo-derivatives the children in the top quartile actually made substantially more errors than the children in the bottom quartile.

The theory to explain the behavior of the high-ability children is straightforward. It is apparent that they are aggressively applying the word-formation rules of English to hypothecate meanings for unfamiliar letter strings. Consider some meanings that might be constructed: loyalment (a devoted band of followers); conversal (the opposite case); assistity (the state or quality of being helpful).

If an adult were to find fault with children who say they know the meanings of pseudo-derivatives, it would be that these forms are not really words in English. But this complaint is based on too narrow a view of the language and overlooks the considerable generative power of morphology. Every day new words are coined that are understood perfectly upon first being used. Probably individual language users employ word-formation processes to produce or understand forms that are not already

stored as "separate entries" in their lexicons, though this is a matter of some debate (Chomsky, 1971; Stanners, Neiser, Hernon, & Hall, 1973).

A more subtle complaint, one that might be raised by a linguist, is that children who call pseudo-derivatives words are failing to acknowledge the blocking or preemption rule (Aronoff, 1976; Clark, in press). This rule says you can't form a new word that means the same thing as an existing word. For example, forgivity is preempted by forgiveness, as long as the two are construed to mean the same thing. But how is one to know in advance that the new form does not differ in some shade of meaning? One does not reject observance because one already knows observation, even though both are nominalizations of observe, and it would be difficult to say exactly what the distinction between them might be on the basis of morphology alone.

In our judgment, knowledge of word formation processes is one of the engines driving vocabulary growth. As the case of observance and observation illustrates, though, the morphology of words may contribute to understanding without providing enough information to precisely determine meaning. Exact distinctions must be resolved in context. Context does not ordinarily provide sufficient clues to determine meaning, either. Together, however, the two sources of incomplete information--morphology and context--may complement one another, so that in combination they provide enough information to pinpoint the meaning of a new word. In the best circumstances, using both morphology and context, it may be possible to learn the meaning of a word in a single encounter. For instance, it is not obvious from morphology what meaning one of our

pseudo-derivatives, observement, should have in relation to observation and observance. Now notice what happens when a context is provided: "The sentry paced back and forth on the observement." At this point it is clear that, if it were a word, observement would refer to a vantage point such as a watch tower.

The fifth graders in the bottom quartile also showed a false alarm rate on pseudo-derivatives that was higher than their average false alarm rate of 29.2%. This suggests that like their high ability cohorts, low ability children are trying to use morphology to figure out the meanings of words. However, the most noteworthy aspect of the performance of the low ability children was their pronounced tendency to false alarm on items that are phonemically or visually similar to real words. Thus, the data provided still another confirmation of the dismal fact that a great many poor readers are also poor decoders.

The good news is that there was an illuminating pattern to the false alarms of the low ability children. The data suggest to us that if these children's first attempt to decode an item matches a word they know, fine. If not, since they recognize they are not very good decoders, they keep jiggering the decoding until they find a match with a known word, or until they run out of decoding options or give up. This theory is diagrammed in Figure 1.

Insert Figure 1 about here.

Furthermore, the false alarm data leads us to conjecture that the typical poor reader tries decoding options in a predictable sequence as follows:

1. Decode the item in the manner preferred in English, or at least in a manner legal in English. Say yes even though the item does not have conventional English spelling. Example: jerbal → gerbil.

2. Change the vowel from short to long, or long to short. Examples: cobe → cob; ritter → writer.

3. Change vowel to a phonemically or visually similar one. Examples: robbit → rabbit or robot; grell → grill.

4. Try another permissible rendering of a consonant. Example: risent → recent.

5. Change a consonant to a phonemically or visually similar one. Examples: blint → blind; flane → flame.

As a partial check on the model, the 25 nonwords least affirmed by the low ability students were examined. If the model is correct, few of these should be transformable into common words using the five rules. In fact, only one could be changed to produce a fairly common word by applying just one rule (sturve → starve). Five more resulted in rare words when a single vowel was changed (ollure, vositation, flort, roversal, munifestation). The remainder required two vowel changes or two consonant changes and the resulting words generally were not common ones. There were, in addition, two pseudo-derivatives (arousion, offendation). Apart from sturve, the items in this least-affirmed list are consistent with the model.

Of course, the foregoing model gives only a partial account of possible transformations poor readers might tinker with when unfamiliar words are encountered. It is partial in both its breadth and depth.

Additional nonwords of different sorts would be required to identify other possible transformation rules.

We would expect that the degree of similarity of the vowels or consonants in a nonword to a related real word would be directly related to the probability of a yes response on the item. That is to say, vowels and consonants can be thought of as having various distances from one another in phonetic space, and we would expect transformations involving neighbors to be more commonly used than those involving far-flung acquaintances. Phoneticians have found it useful to use a spatial representation, as in Figure 2, to chart the production of sounds in the mouth. Of course, the location of particular sounds varies among accents and speakers. Nonetheless, it may be possible to make predictions of false alarm rates for particular nonsense words based on the distance to be traveled on the vowel chart before a familiar meaningful word is produced. A complete theory of false alarms would also have to take account of phonetic similarity among consonants, graphic similarity, and probably other sources of confusability. In the meantime, the general point is that false alarm patterns based on recoding distance of nonsense forms to meaningful words might prove valuable as a diagnostic tool for the language teacher, serving to pinpoint the areas in which knowledge of sound-to-letter correspondences are weakest.

Insert Figure 2 about here.

If the model that has been proposed to explain the false alarms of poor decoders is on the right track, then poor decoders may also be expected to produce a certain number of mock hits. A "mock hit" can be

defined as saying yes to an unknown word as the result of having transformed it into a known one. For instance, the following is an especially likely mock hit: sham → shame. Mock hits will inflate the scores of poor readers, and they pose a treacherous problem in estimating vocabulary size (see next section) because they distort the function that relates the probability that a word is known to frequency of usage.

Poor decoders are also vulnerable to "incorrect rejections," that is, saying no to known words that have been misdecoded. To express this fact in traditional terms, a poor decoder's reading vocabulary is not as large as his or her listening vocabulary.

We believe that the phonemic experimentation apparently engaged in by low ability students can be thought of profitably as a hierarchy, or "stack," of transformations arranged in order of amount of deviation from spelling-to-sound conventions. Our notion is similar in conception to the transformation stack system devised by Prytulak (1971) to account for the elaborations people invent when trying to remember lists of nonsense syllables. When a person attempts to learn a nonsense syllable, Prytulak argued, he or she seems to work down through an ordered list of transformation options, trying one option after another, until a meaningful representation can be generated.

The concept of stack depth may have some heuristic value in getting beyond the notion that poor decoding consists of a miscellaneous jumble of mistakes. It could be that students vary systematically in the depth they will go in order to recode a letter string into a familiar word. For instance, one child might freely interchange long and short vowels

but completely avoid grosser transfigurations such as risent → recent, whereas another might make both kinds of substitutions.

Following Brown and Burton (1978), another possibility is that children have particular kinds of "bugs" in their decoding procedures. For example, a certain child might have a propensity to switch short and long vowels and terminal ts and ds, producing false alarms like blint → blind. Other children would produce characteristic false alarms of a different sort, depending upon their particular bugs. Kubato (1981) has explored the possibility that fifth graders in our sample had systematic bugs in their decoding routines. The conclusion was that the approach was promising; however, that before the promise could be realized it would be necessary to more systematically vary the features of the nonwords, in such a fashion that hypothesized bugs could be reliably identified. This work has not been undertaken as yet.

In summary, the yes/no test shows considerable promise as an inexpensive diagnostic tool. It should be cautioned, however, that the ideas presented in this section are speculative. We have not even taken such obvious steps as seeing whether children can come up with reasonable meanings for pseudo-derivatives they think they know or whether they will pronounce nonwords in accordance with hypothesized phonemic and graphic transformations.

Estimating Absolute Vocabulary Size

Our original hope when we began to investigate the yes/no task was that we would be able to develop a simple yet accurate method of estimating the number of words a person knows (Anderson & Freebody, 1981). For a

variety of reasons, we have not reached that goal so far. This section reports our progress to date.

The critical problem is how to get a precise estimate of vocabulary knowledge, separate from tendency to over- or underestimate this knowledge. The high threshold model is one solution, but probably not the best that can be devised. The problem with the high threshold model is that it does not accommodate gracefully to degrees of knowing short of perfect knowledge. Both theory, and data that we have gathered but will not report here, indicate that knowledge of word meanings is seldom all or none, that a person can know some of the distinctions conveyed by a word without knowing them all. For instance, one could know that tort is a legal term without knowing exactly what it means.

In collaboration with Michael Levine, we have been developing a latent trait, logistic model that we believe will be an improvement on the high threshold model. According to the current version of our model, individual readers are ordered according to overall word knowledge, $\underline{\theta}$, and judgmental standards, or degree of conservatism, $\underline{\beta}$. The "depth" of word knowledge for a certain person at level $\underline{\theta}$ on the $\underline{i}$th item in the list of words and nonwords is:

$$\underline{a_i}\ \underline{\theta} + E$$

Here $\underline{a_i}$ is a parameter quantifying word properties such as frequency of use and $\underline{E}$ is a random variable with a bell shaped density. If depth exceeds criterion β the person responds affirmatively. Thus the conditional probability of a yes response is Prob $\{\underline{a_i}\ \underline{\theta} + \underline{E} \geq \underline{\beta}\}$. The parameter $\underline{a_i}$ is positive for words, negative for most nonwords, close to

zero for very hard items, and large in absolute value for easy items. The person parameter $\underline{\beta}$ is large for conservative readers and small, perhaps negative, for less conservative readers. The person parameter $\underline{\theta}$ is positive and large for able readers who consistently distinguish words and nonwords and positive but smaller for less able examinees.

The parameters, $\underline{a}$, $\underline{\theta}$, $\underline{\beta}$, are determined by maximum likelihood estimation. At this time we have developed numerically stable parameter estimation computer programs. These programs have delivered reliable estimates of person parameters in the preliminary study with fifth grade students. With a large enough sample of words, the parameters of word knowledge, $\underline{\theta}$, and conservatism, $\underline{\beta}$, can be estimated to within any specified margin of error for any individual.

The distribution of words in the language according to frequency is known to be log normal (Carroll, Davies, & Richman, 1971). We propose to take advantage of this fact in estimating vocabulary size. Our early results suggest that the function relating $\underline{a}$ in our model to frequency of usage is very regular. If these results hold up, estimating vocabulary size will simply be a matter of integrating under the function. Indeed, we have already made trial estimates for our sample of fifth graders that look quite sensible. These estimates are shown in Figure 3. The scales in this figure were deliberately made grainy, since the actual values should not be taken seriously: the children are not a random sample of fifth graders; the words are not representative of school English (too few very infrequent words); and the raw data were smoothed to make the curves look nice.

Insert Figure 3 about here.

We anticipate that our approach eventually will permit a number of reliable and useful statistics about breadth and depth of vocabulary knowledge. Tables or graphs could be prepared for each grade showing the total number of distinct word families known by children at benchmark percentile ranks among their grade cohorts, perhaps the 90%-tile, 70%-tile, 50%-tile, 30%-tile, and 10%-tile, as in Figure 3. Instead of number of word families known, the statistic could be the proportion of words known from the most frequent 4,000, 10,000, or 30,000 words in the language. Alternately, the statistic could be the estimated number of words that a child would know in 1,000 running words of reading material. Since the model is expected to be able to predict depth of word knowledge, the choice of a strict or lenient standard of what it means to "know" a word can be made with respect to any statistic that might be devised. The theory for tailoring a test to the individual is especially simple (assuming an unidimensional item pool).

In order to attain the goal of absolute estimates of number of words known, several steps will have to be completed. First, the model for disentangling word knowledge from judgmental standards will have to be perfected and thoroughly evaluated. Second, a procedure will have to be devised for drawing samples of words stratified according to frequency of usage which takes account of the fact that the standard error of estimate of frequency increases as frequency decreases. Third, generous samples of words and nonwords will have to be given to people of various ages to provide normative data.

The Relation of Vocabulary Knowledge to Reading Comprehension

Vocabulary difficulty has always proved to be a factor of overpowering importance in studies of readability. Thus, it is most surprising that experiments that have directly manipulated word difficulty and tested the effects on comprehension have produced weak, conflicting results. Marks, Doctorow, and Wittrock (1974), and Wittrock, Marks, and Doctorow (1975) have reported that replacement of about 15% of words in a passage with very rare synonyms resulted in significant decreases in reading comprehension. Three studies, on the other hand, have found that explicit, demonstrably-successful instruction in vocabulary fails to increase students' comprehension of texts containing the taught words (Tuinman & Brady, 1974; Pany & Jenkins, 1977; Jenkins, Pany, & Schreck, 1978).

Many differences between the materials and procedures of these studies and those employed by Marks and her associates might account for the discrepant findings. Among these could be length of passages, degree of difficulty of the words, the measures of comprehension used, and so on. We will summarize here a program of research in which we are engaged that is attempting to clarify the role of vocabulary knowledge in text comprehension. Specifically, we have attempted to answer the following four questions: (a) What proportion of the substance words in a text need to be unfamiliar before comprehension shows reliable decreases? (b) Does the effect of vocabulary difficulty depend upon whether the unfamiliar words are located in important or unimportant ideas in the text? (c) Does the effect of vocabulary difficulty depend upon the cohesiveness of the text? (d) Does the effect of vocabulary difficulty

depend upon whether the reader has available a familiar schema to assimilate the text?

In this series of experiments, reported fully elsewhere (Freebody & Anderson, 1981a, 1981b), the passages were about 300 words in length. They were selected from Scott Foresman Social Studies for fifth grade, except for those in one of the studies which were written at a similar level. The measures of comprehension were free recall, summarization, and true/false sentence verification. The subjects were sixth-grade students ranging from below average to well above average in language ability. The students were tested in their intact class groups.

In the first experiment, we examined the issue of the proportion of rare words in a passage that could be substituted in a text before comprehension suffered. Seventy-two sixth graders read three social studies passages. For each student, one passage had easy vocabulary; one was medium in difficulty, in which one substance word in six was changed to a rare synonym; and one had difficult vocabulary, in which one substance word in three was a rare synonym for the original.

We found a significant effect on only one measure, the sentence verification test. On the recall measure, there was a trend toward better performance when the vocabulary was easy; for 8 out of 9 passages the mean recall was higher in the easy form than the difficult form. The effects of medium vocabulary difficulty were inconsistent.

The answer to the first question is that a rather high proportion of unfamiliar vocabulary is required before a consistent decrease in performance results. Roughly half of the words in any passage are substance words. Thus, in a 300 word passage there are about 150 substance

words and 50 of them had to be changed to rare synonyms before there was a discernible effect. This seems to us to be a strikingly high proportion.

Does it matter where difficult vocabulary appears? It seems reasonable to suggest that, in an extreme case, one unfamiliar word could render an otherwise simple passage incomprehensible. Similarly it may be that, if the important ideas in a passage are accessible, a very high proportion of unknown words in the other sections of text will not matter. We had sixth grade students rate each proposition in three passages for importance. Thus we had a mean importance ranking for each proposition in each passage. The most important and least important fourths of the propositions were identified in order to produce three forms of each of three social studies passages: an easy form with high frequency words only, a difficult-unimportant form in which at least one rare substitution was included in each of the least important propositions, and a difficult-important form containing rare synonyms for the original words in each of the most important propositions. This technique produced a proportion of rare words in the latter two passages of about one in nine. As in the first experiment, each student read a passage in each vocabulary form, with order and passage counterbalanced. Of major interest to us was whether the location of unfamiliar vocabulary in important or unimportant propositions in a text made a difference to comprehension.

The most noteworthy finding of the experiment was that passages containing unfamiliar vocabulary in unimportant propositions were significantly better summarized than passages containing unfamiliar vocabulary in important propositions. Our conjecture is that when a

reader encounters unfamiliar words he or she often does not completely process the proposition containing them. This leaves fewer propositions to be processed and results in better encoding or greater accessibility of the remaining propositions. Therefore, when it is the unimportant propositions that contain hard words, the important ones are readily available for inclusion in a summary.

In this experiment, the results on the recall and sentence verification measures were unclear because of hard to interpret interactions.

Our third experimental question was: Does text cohesion interact with vocabulary difficulty to diminish the negative effects of unfamiliar vocabulary on comprehension? Information is repeated, explicitly, in most texts, and this redundancy may permit the reader either to ignore unfamiliar words and search elsewhere for sufficient clues to meaning to allow fluent processing to continue, or even to use the context to determine a rare word's meaning. These clues will be both semantic and syntactic, and will be available and unambiguous to the degree that the text is cohesive.

Haliday and Hasan (1976) have identified five types of linguistic cohesion in text: (a) reference, in which an element needs, for its interpretation, to be related to another thing, class of things, place, or time, (b) substitution, where an element is replaced by another term, (c) ellipsis, in which an element is omitted but understood, (d) conjunction, and (e) lexical cohesion, in which an element is either repeated or replaced by a synonym, a superordinate, a general word, or in which a "collocation" has occurred--that is, in which lexical items are used which regularly co-occur. When cohesion is high, the reader

presumably can easily retrieve relevant information and integrate it into the new proposition. The clues to do this may be a referential, substitutive or elliptic device, but the operation seems essentially the same.

Using this taxonomy, "low cohesiveness" can be operationalized as the downgrading of referential, substitutive, and elliptic devices and by infrequent conjunction. Ties may be arranged hierarchically in terms of the burden they impose on processing. Repetition of a referential term may be supposed to entail the least processing effort, followed by common synonym substitution, pronominalization, and ellipsis. In order to make a text less cohesive in these terms, a tie would need to be replaced by a tie at least one step lower in this hierarchy.

The following excerpts illustrate the high and low cohesion passages used in the study.

High cohesion

All countries have laws about how trade and business can be carried on with other countries. One of the oldest ways that governments control trade with these laws is through a "tariff" law. The tariff is most often a tax on goods coming into a country. The tax is added to the goods and so it makes the goods cost more.

Low cohesion

All countries have laws about how trade and business can be carried on with other countries. One of the oldest ways that governments control exchange is through a "tariff" law. This is most often a tax on goods coming into a country. It is added to their price and so makes them cost more.

More gross disruptions to text cohesion are possible. An author, for instance, may fail to reiterate an earlier stated proposition which is important for an understanding of the discourse at hand. Implicit or unpredictable premises may be used to link new topics, and extraneous information may be gratuitously included. These have been called instances of "inconsiderateness" (Kantor, 1978). Here is an example of inconsiderateness taken from the passage describing the nature and purpose of tariff laws: Following the statement that luxuries such as furs and perfumes are the objects of particularly severe tariffs, there is a sentence to the effect that France has always been famous for popular perfumes. A referential tie exists (the repetition of "perfumes"), and a weak lexical collocation could be in effect since trade has presumably been discussed in terms of imports from other countries and "France" is a member of the category "other countries." So superficially the sentence is adequately tied. However, the reader is led to process extraneous information, which perhaps causes fruitless searches of memory, or which causes the development of unfulfilled expectations. Irrelevant material in the text would, it is hypothesized, place additional burdens on the reader, and hamper the development of ideas about the meanings of text segments containing unfamiliar words.

To summarize, three levels of cohesion were developed for each passage used in this experiment--high, low, and inconsiderate. Highly cohesive passages contained frequent referential repetition, synonymy, and conjunction. In the low cohesion forms, the ties were downgraded to produce more pronominalization and ellipsis, and many conjunctions were removed. To produce the inconsiderate forms, eight extraneous

propositions were added at four equally spaced intervals to the low cohesion forms of the passages. Each of these three cohesion conditions appeared in two vocabulary conditions, easy and difficult. The difficult vocabulary versions were produced by substituting a rare synonym for one substance word in four. Each of 75 sixth grade students read three passages, one in each cohesion condition. Half the students read passages with easy vocabulary and half with difficult.

The major issue was whether the effects of unfamiliar vocabulary on the three measures of comprehension depend upon the degree of linguistic cohesion in a text. Specifically, we hypothesized that differences between vocabulary levels would be minimal when the text was highly cohesive, but more considerable as the cohesion diminished. This prediction was not confirmed. While there were effects for vocabulary difficulty on the recall and summarization measures, there was no interaction between vocabulary difficulty and cohesion level. There was an interaction between cohesion level and order in which the passage was read: High cohesion was associated with better free recall when a highly cohesive passage is read first, while inconsiderateness and low cohesion depressed performance when those conditions are encountered later. The interaction between cohesion level and the order of reading suggests reader fatigue in the processing of cohesive devices. Perhaps, as the reader becomes tired or loses interest, one of the processes that suffers is the making of linking inferences, such as finding pronouns' coreferents, making conjunctive links, and so on.

The fourth question: Does schema availability interact with vocabulary difficulty such that when a familiar schema is available

unfamiliar vocabulary is less detrimental? To answer this question we selected two themes, a game theme and a visit theme. For each theme there was a certain script. For the game theme, for instance, the script dealt with the inventors of the game, the objects used, the terrain needed, the grips preferred, and the climate required. Based on each theme we wrote, in sentences identical in their syntactic structure, two passages--a familiar instantiation of the theme and an unfamiliar instantiation. For the game theme, the two passages dealt with a game of horseshoes as played by cowboys and a game called "Huta" played by American Indians with a buffalo bone. The visit script was instantiated, first, as a visit to a supermarket and, second, as a trip to a Niugini Sing-Sing, an intertribal musical get-together. Each of these four passages also appeared in two vocabulary levels, easy and difficult. One substance word in four was changed to a rare synonym to produce the difficult vocabulary versions. Only those substance words common to both the familiar and unfamiliar versions were changed.

We want to emphasize the high degree of control we gained over extraneous factors in this experiment. The sentences in familiar and unfamiliar versions of a theme were identical in their syntactic structure, and many of the words were common. An example will give the flavor of the contrast. The two following passages are the opening excerpts from the familiar and unfamiliar passages instantiating the visit theme:

Supermarkets

I once got to be the friend of a family who lived in the jungles of Niugini. While I was staying with them once, I happened to say that their food was much tastier than the food we Americans bought in our supermarkets. "Your what?" they asked. They had never heard of supermarkets.

Niugini Sing-Sing

I once got to be the friend of a family who lived in the jungles of Niugini. While they were staying with me once, they happened to say that our music was much noisier than the music they made in their sing-sings. "Your what?" I asked. I had never heard of sing-sings.

Obviously some changes in vocabulary were necessary but nonetheless it can be seen that the match was close and the distinct vocabulary in the two versions was matched in terms of length and frequency.

There were 82 sixth-grade students in this study. Each student read the familiar passage for one theme and the unfamiliar form for the other. Half the students were in the difficult vocabulary condition and half in the easy condition. As in the previous experiment, our major interest was in the interaction, in this case between vocabulary difficulty and schema availability. This interaction was not a significant effect for any of the comprehension measures. Vocabulary difficulty made a difference on the sentence verification task, and there was a trend on the free recall task. There were no clear findings involving the summarization measure. Essentially, for recall and sentence verification, both vocabulary difficulty and familiarity affected performance, but there was no lessening of the vocabulary effect in the

familiar condition, nor was there a severe depression of performance for the unfamiliar topic and rare vocabulary forms of the passages.

We have summarized four studies, which made up our initial attempts to examine the effects of vocabulary difficulty on reading comprehension, and its possible interaction with high-order text factors. We now wish to draw some overall conclusion about the effects of including rare words in a text on students' comprehension. For all three measures in each of the four experiments, vocabulary difficulty effects, while not all significant, were always in the expected direction. That is, rare words always tended to lead to lower performance. An effect-size analysis (McGaw & Glass, 1980) was conducted to describe the overall impact of difficult versus easy vocabulary in standard deviation units. The mean effect size for recall was 2.7, for summarization the mean was 1.4, and for sentence verification, it was 2.0. These may be interpreted as indicating that the comprehension performance of the 50th percentile student reading a passage with easy vocabulary would cause that student to be ranked, among an equivalent group reading that passage with difficult vocabulary substitutions, at the 99th percentile on recall, the 93rd percentile on summarization, and the 96th percentile on sentence verification. Over all measures, the mean effect size was 2.1, an overall performance equivalent to the 98th percentile.

It can be asserted with some confidence, then, that vocabulary difficulty, as defined in these experiments, is related to measures of text comprehension. At the same time, it should be noted again that a large proportion of words have to be changed in order to see reliable

effects, and it should be emphasized that the effects of hard words were never very large in absolute terms.

The failure of level of vocabulary difficulty to interact with either text cohesion or schema availability is surprising. The view that reading is an interactive process is now widely accepted among reading researchers. In essence, the theory says that information from many levels of analysis is integrated during reading. A corollary is that if information from one level is unavailable, the reader will generally be able to compensate by using information from other levels. There was no evidence to support the compensation hypothesis in the experiments summarized here.

Conclusions

Our most important finding is about assessment. The yes/no test has great promise for broad-gauged measurement of knowledge of word meanings. A yes/no vocabulary test is simple to construct and simple to calibrate. An item for a yes/no test is simply a word or nonword letter string. It is not embedded in a complex context of distractors constructed with reference to a specific age group. There is no need for trained item writers or a secure item pool.

The directions for a yes/no test are readily understood by first graders. The yes/no test minimizes extraneous demands for a strategic knowledge or ability in self-expression.

A yes/no test makes efficient use of time; over twice as many words can be examined in an interval of time on a yes/no test than on a multiple choice test.

Most important, a score on a yes/no test provides a much more valid indicator of whether an examinee actually knows the meanings of the tested words than a score on a standardized multiple choice test.

Even a simple high-threshold correction of yes/no scores does passably well at separating word knowledge from the tendency to over- or understate this knowledge, and we believe we are within reach of a superior model for disentangling the two facets of performance. If this goal is reached, it should prove possible to make accurate estimates of the number of words a child knows.

On the negative, a yes/no test is unsuitable for determining whether a person knows a particular meaning of a word with many meanings. It is also unsuitable for evaluating the effects of direct vocabulary instruction since an examinee would be able to recognize that a word is familiar without knowing its meaning. Indeed, a possible general problem with the yes/no method is that it will not satisfactorily distinguish between knowledge of meanings and mere familiarity.

The false alarms (saying yes to nonwords) that children make on a yes/no test provide interesting insights into their language processes. All fifth graders, but most especially good readers at this level, false alarm on pseudo-derivatives such as loyalment and adjustion. This indicates aggressive application of morphological principles to attack the meanings of unfamiliar words.

Analysis of false alarms suggests that poor fifth grade readers, and only the poor readers, engage in phonemic experimentation with unfamiliar items in order to try to find a match with words they know.

That is to say, for instance, if a poor reader cannot match a known word by giving the main vowel a short sound he or she may try giving it a long sound, whether or not the spelling-to-sound rules of English permit a long sound in that context. An exciting possibility is that a properly designed yes/no test may yield, as a by-product, a profile of the "bugs" in a child's decoding procedures.

Four experiments were summarized which show beyond any reasonable doubt that vocabulary difficulty does influence text comprehension, though the effects of difficulty were not as strong as one might expect on the basis of readability research. Some subtle effects of hard words were uncovered. One of these is that when the hard words appear only in unimportant propositions, students' summaries of texts actually improve. Another is that vocabulary difficulty does not interact with either text cohesion or schema availability, a result which is puzzling when looked at from the perspective of an interactive theory of reading.

Reference Note

1. Nagy, W. E., & Anderson, R. C. The number of words in printed school English (Tech. Rep.). Champaign: University of Illinois, Center for the Study of Reading, in press.

References

Anderson, R. C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42, 145-170.

Anderson, R. C., & Freebody, P. Vocabulary knowledge (Tech. Rep. No. 136). Urbana: University of Illinois, Center for the Study of Reading, August 1979. (ERIC Document Reproduction Service No. ED 177 480)

Anderson, R. C., & Freebody, P. Vocabulary knowledge. In J. T. Guthrie (Ed.), Comprehension and teaching: Research reviews. Newark, Del.: International Reading Association, 1981.

Aronoff, M. Word formation in generative grammar. Cambridge, Mass.: M.I.T. Press, 1976.

Brown, J. S., & Burton, R. R. Diagnostic models for procedural bugs in basic mathematic skills. Cognitive Science, 1978, 2, 155-192.

Carroll, J. B., Davies, P., & Richman, B. American Heritage: Word frequency book. Boston, Mass.: Houghton Mifflin, 1971.

Chomsky, N. Remarks on nominalization. In N. Chomsky (Ed.), Studies on semantics in generative grammar. The Hague: Mouton, 1971.

Clark, E. V. The young word-maker: A case study of innovation in the child's lexicon. In L. Gleitman & E. Wanner (Eds.), Language acquisition: The state of the art. Cambridge: Cambridge University Press, in press.

Freebody, P., & Anderson, R. C. Effects of differing proportions and locations of difficult vocabulary on text comprehension (Tech. Rep. No. 202). Urbana: University of Illinois, Center for the Study of Reading, May 1981. (ERIC Document Reproduction Service No. ED 201 992) (a)

Freebody, P., & Anderson, R. C. Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension (Tech. Rep. No. 225). Urbana: University of Illinois, Center for the Study of Reading, November 1981. (ERIC Document Reproduction Service No. ED 170 788) (b)

Haliday, M. A. K., & Hasan, R. Cohesion in English. London: Longmans, 1976.

Hall, W. S., & Tirre, W. C. The communicative environment of young children: Social class, ethnic and situational differences (Tech. Rep. No. 125). Urbana: University of Illinois, Center for the Study of Reading, May 1979. (ERIC Document Reproduction Service No. ED 170 788)

Jenkins, J. R., Pany, D., & Schreck, J. Vocabulary and reading comprehension: Instructional effects (Tech. Rep. No. 100). Urbana: University of Illinois, Center for the Study of Reading, August 1978. (ERIC Document Reproduction Service No. ED 160 999)

Kantor, R. N. Anomaly, inconsiderateness, and linguistic competence. In D. Lance & D. Gulstad (Eds.), Papers from the 1977 Mid-America Linguistic Conference. Columbia: University of Missouri, 1978.

Kubato, K. Towards an error analysis in vocabulary acquisition. Urbana: University of Illinois. Unpublished masters thesis, 1981.

Marks, C. B., Doctorow, M. J., & Wittrock, M. C. Word frequency and reading comprehension. Journal of Educational Research, 1974, 67, 259-262.

McGaw, B., & Glass, G. V. Choice of the metric for effect size in meta-analysis. American Educational Research Journal, 1980, 3, 325-337.

Miller, G. A. The acquisition of word meaning. Child Development, 1978, 49, 999-1004.

Mills, C. W. The Sociological Imagination, Pelican, London, 1970.

Pany, D., & Jenkins, J. R. Learning word meanings: A comparison of instructional procedures and effects on measures of reading comprehension with learning disabled students (Tech. Rep. No. 25). Urbana: University of Illinois, Center for the Study of Reading, March 1977. (ERIC Document Reproduction Service No. ED 136 237)

Parsons, T. The Social System, Glencoe, Illinois, Free Press, 1951.

Prytulak, L. Natural language mediation. Cognitive Psychology, 1971, 2, 1-56.

Sims, V. M. The reliability and validity of four types of vocabulary test. Journal of Educational Research, 1929, 20, 91-96.

Stanford Achievement Tests, N.Y.: Harcourt, Brace, Jovanovich, 1973.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. Memory representation for morphologically related words. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 399-412.

Terman, L. M. Vocabulary tests as a measure of intelligence. Journal of Educational Psychology, 1918, 9, 452-466.

Thorndike, R. L. Reading comprehension education in fifteen countries. N.Y.: Wiley, 1973.

Tuinman, J. J., & Brady, M. E. How does vocabulary account for variance on reading comprehension tests? A preliminary instructional analysis. In P. Nacke (Ed.), Twenty-third yearbook of the National Reading Conference. Clemson, S.C.: National Reading Conference, 1974.

Wittgenstein, L. Tractatus Logico-Philosophicus. London: Routledge & Kegan Paul, 1961.

Wittrock, M. C., Marks, C., & Doctorow, M. Reading as a generative process. Journal of Educational Psychology, 1975, 67, 484-489.

Zimmerman, J., Broder, P. K., Shaughnessy, J. J., & Underwood, B. J. A recognition test of vocabulary using Signal-Detection measures, and some correlates of word and nonword recognition. Intelligence, 1977, 1, 5-31.

Table 1

Most Frequent False Alarms of Low and High Ability Students

(with other group's percentage in parentheses)

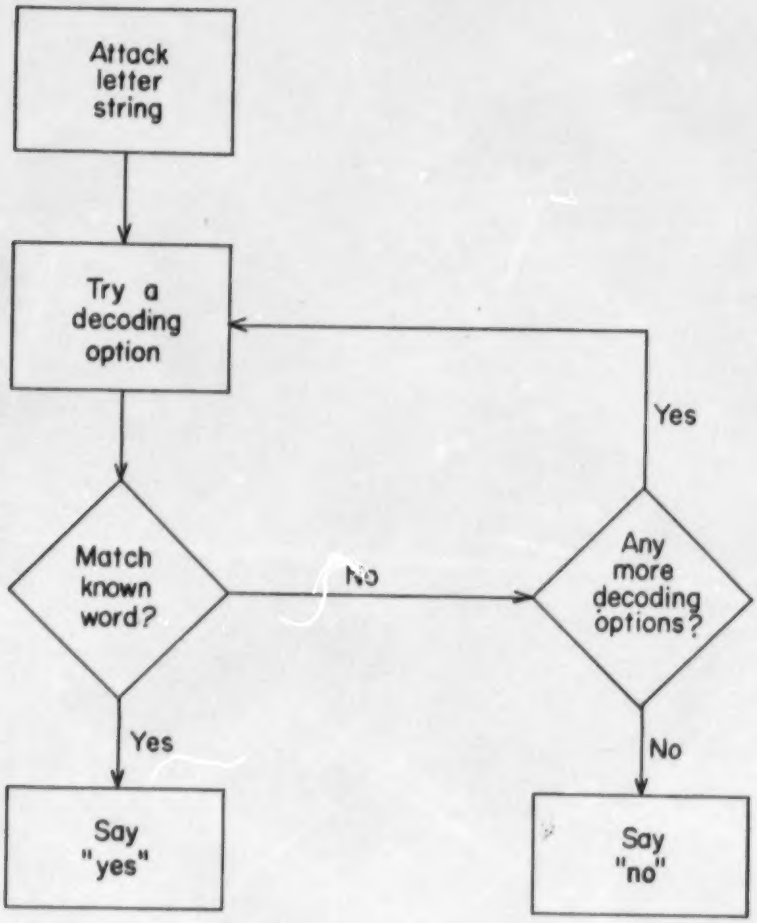
Low Ability			High Ability		
Nonword	Percentage		Nonword	Percentage	
jerbal	67	(19)	loyalment	70	(44)
cobe	59	(0)	successment	67	(48)
bighter	56	(4)	observement	59	(41)
robbit	56	(0)	conversal	48	(40)
slead	52	(15)	adjustion	37	(37)
porfame	52	(4)	deformness	33	(44)
flane	52	(0)	assistity	33	(19)
successment	48	(67)	instructness	30	(33)
risent	48	(19)	persistion	26	(37)
mudge	48	(11)	jerbal	19	(67)
compure	48	(7)	risent	19	(48)
plode	48	(4)	issuance	19	(30)
revese	48	(4)	forgivity	19	(19)
breat	48	(0)	rehearsion	19	(4)
grell	48	(0)	slead	15	(52)
weast	48	(0)	arousion	15	(11)
loyalment	44	(70)			
deformness	44	(33)			
lote	44	(11)			
strangity	44	(11)			
ritter	44	(7)			
blint	44	(4)			
sleem	44	(4)			
bleen	44	(0)			
pless	44	(0)			

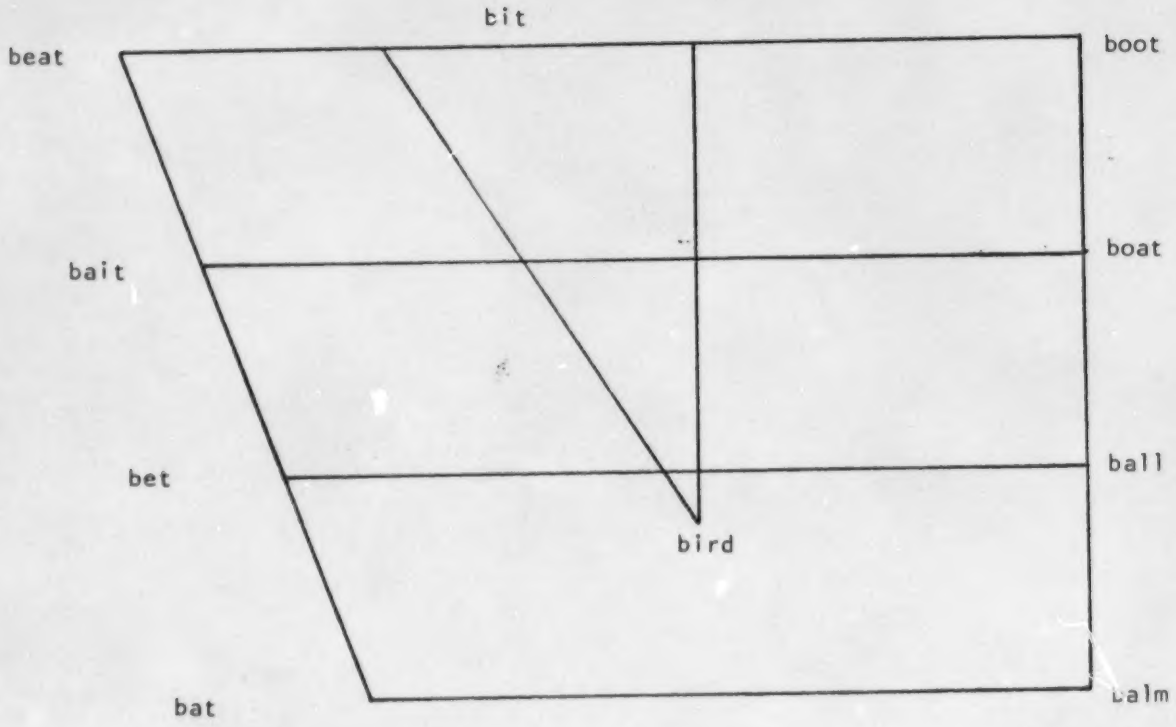
Figure Captions

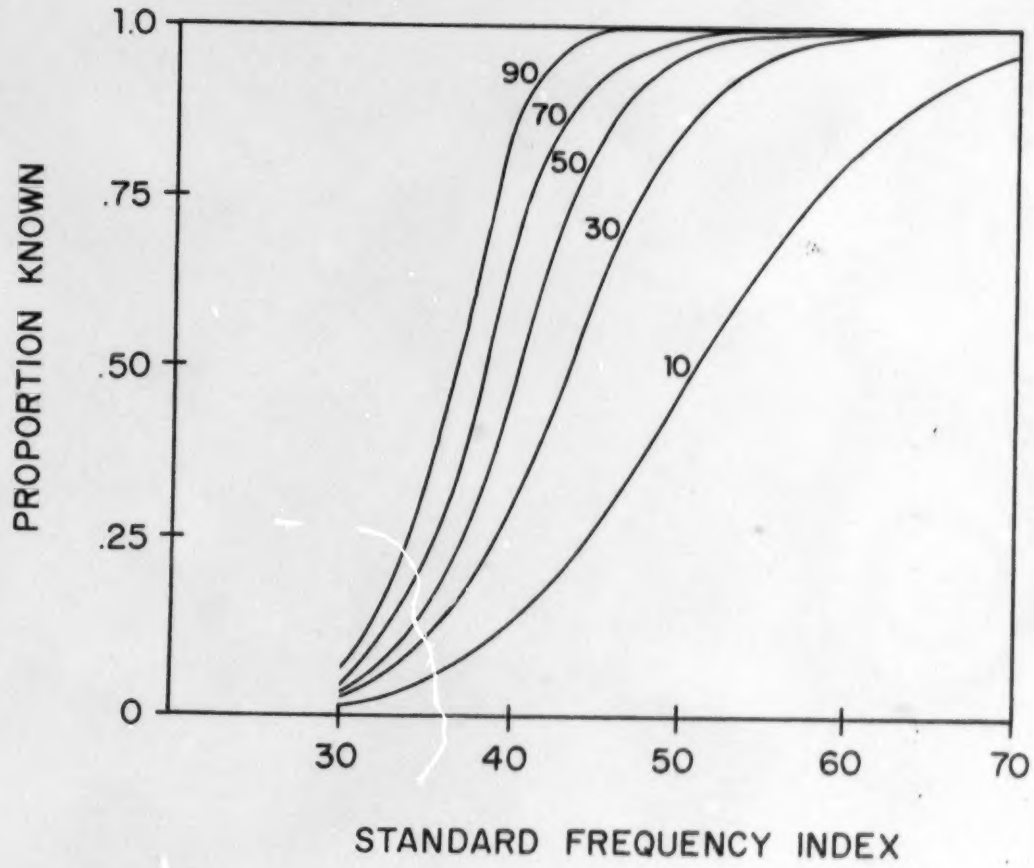
Figure 1. Hypothetical decoding strategy of poor reader on yes/no task.

Figure 2. A cardinal vowel chart showing some examples of vowel locations.

Figure 3. Best-fitting functions for the relationship between knowledge and word frequency for five percentile groups.


Attack letter string
Try a decoding option
Match known word?
No
Any more decoding options?
Yes
Yes
No
Say "yes"
Say "no"


bit
beat
boot
bait
boat
bet
ball
bird
bat
balm


1.0
.75
.50
.25
0
PROPORTION KNOWN
90
70
50
30
10
30
40
50
60
70
STANDARD FREQUENCY INDEX